

Efficiency of text readability features in Russian academic texts

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2018 Rossiiskii Gosudarstvennyi Gumanitarnyi Universitet. All Rights Reserved. This paper addresses the problem of readability assessment for Russian texts and investigates the impact of 24 lexical, syntactic and frequency features. The research was conducted on Russian Readability Corpus containing two sub-corpora, two sets of 5–11 grade level textbooks on Social studies for native speakers of Russian. The sub-corpora were collected for research purposes, annotated and marked as BOG and NIK. The application of the Ridge regression has demonstrated the connection between readability and average sentence length, average number of coordinating chains, average number of sub-trees, frequency and lexical features. The results of the study have the potential to be applied in a wide variety of areas including primarily education, as well as webpage design, document management.

Keywords

Average number of coordinating chains, Average sentence length, Readability assessment, Russian Readability Corpus

References

- [1] Biber, D. (1990). Methodological Issues Regarding Corpusbased Analyses of Linguistic Variation. *Literary and Linguistic Computing*, 5: 257–69.
- [2] Chall, J. S. (1958). *Readability: An appraisal of research and application*. Bureau of Educational Research Monographs, No. 34. Columbus, Ohio State Univ. Press.
- [3] Chen, X. B., Meurers, D. (2016). Characterizing Text Difficulty with Word Frequencies. In *Proceedings of The 11th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 84–94). San Diego, CA. Association for Computational Linguistics.
- [4] Choldin, M. T. (1979). Rubakin, Nikolai Aleksandrovic. In A. Kent, H. Lancour, J. E. Daily (Ed.) *Encyclopedia of library and information science*. (pp. 178–179). Basel: CRC Press.
- [5] Krioni N. K., Nikin A. D., Filippova A. V. (2008). Avtomatizirovannaya sistema analiza parametrov slozhnosti uchebnogo teksta In *Tekhnologiya i organizatsiya obucheniya: nauch. izdanie*. — Ufa: UGATU, P. 155–161.
- [6] Lerner, I. Ya. (1974) Kriterii slozhnosti nekotorykh elementov uchebnika: Problemyshkol'nogouchebnika [The criteria for the complexity of some elements of the textbook: Problems of a school textbook]. Is. 1. Moscow: Prosveshchenie.
- [7] Mackovskij, M. S. (1976). Problemy chitabel'nosti pechatnogo materiala. Smyslovoe vospriyatie rechevogo soobshcheniya v usloviyah massovoj kommunikacii [Problems of readability of printed material. Semantic perception of speech messages in conditions of mass communication]. Moscow: Nauka.
- [8] Mikk, Y. A. (1970) O faktorakh ponyatnosti uchebnogo teksta [On factors of comprehensibility of educational texts]. Diss. ... cand. ped. sciences. Tartu.

- [9] Mutt, O. (1984). Aktual'nye voprosy otbora uchebnogo materiala dlya vuzovskogo kursa inostrannogo yazyka [Actual questions of selection of the educational material for the university course of a foreign language]. Tartu: The Tartu State University
- [10] Osborneva, I. V. (2006) Avtomatizirovannaya otsenka slozhnosti uchebnykh tekстов na osnove statisticheskikh parametrov [Automated estimation of complexity of educational texts on the basis of statistical parameters]. Pedagogy Cand. Diss. Moscow.
- [11] Sharoff, S., Kurella, S. and Hartley, A. (2008). Seeking needles in the web's haystack: Finding texts suitable for language learners. In Proceedings of the 8th Teaching and Language Corpora Conference (TaLC8).
- [12] Solovyev, V., Ivanov, V., Solnyshkina M. (2018) Assessment of reading difficulty levels in Russian academic texts: Approaches and Metrics. Journal of Intelligent & Fuzzy Systems, (Preprint):1–10, 2018.
- [13] Tomina, Yu. A. (1985) Ob"ektivnaya otsenka yazykovoy trudnosti tekстов (opisanie, povestvovanie, rassuzhdenie, dokazatel'stvo) [An objective assessment of language difficulties of texts (description, narration, reasoning, proof)]. Abstract of Pedagogy Cand. Diss. Moscow.
- [14] Tsetlin, B. C. (1980) Didakticheskie trebovaniya k kriteriyam slozhnosti uchebnogo materiala [Didactic requirements to the complexity criteria of educational material]. Novye issledovaniya v pedagogicheskikh naukakh. 1 (35). pp. 30–33.
- [15] Tuldava, Yu. A. (1975) Ob izmerenii trudnosti tekстов [On measuring the complexity of the text]. Uchenye zapiski Tartuskogo universiteta. Trudy po metodike prepodavaniya inostrannykh yazykov. 345. pp. 102–120.
- [16] Ushakov, K. M. (1980) O kriteriyakh slozhnosti uchebnogo materiala shkol'nykh predmetov [On the criteria of complexity of teaching material of school subjects]. Novye issledovaniya v pedagogicheskikh naukakh. 2 (36). pp. 33–35.
- [17] Wessel N. van Wieringen (2018) Lecture notes on ridge regression. arXiv:1509. 09169v2 [stat.ME] 6 Jan 2018.